UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK
----------x

HACHETTE BOOK GROUP, INC., et al.,

Plaintiffs,

v. 20 Civ. 4160 (JGK)(OTW)

INTERNET ARCHIVE,

Defendant.

Conference
----------x

New York, N.Y.
December 2, 2021
4:15 p.m.

Before:

HON. ONA T. WANG,

U.S. Magistrate Judge

APPEARANCES

DAVIS WRIGHT TREMAINE LLP
Attorneys for Plaintiffs
BY: ELIZABETH A. McNAMARA
-and-
OPPENHEIM & ZEBRAK, LLP
BY: SCOTT A. ZEBRAK

DURIE TANGRI LLP
Attorneys for Defendant
BY: JOSEPH C. GRATZ

(Case called)

THE DEPUTY CLERK: Counsel, please state your appearances for the record.

MS. McNAMARA: Yes, Good afternoon, your Honor. Elizabeth McNamara of Davis Wright Tremaine, for the plaintiffs.

MR. ZEBRAK: Good afternoon, your Honor. Scott Zebrak, also for the plaintiffs, Oppenheim & Zebrak LLP.

MR. GRATZ: Good afternoon, your Honor. Joseph Gratz from Durie Tangri in San Francisco, on behalf of defendant Internet Archive.

THE COURT: We've had so many travelers come in today.

All right. Welcome. As you can see, we are still observing full masking protocols in the courtroom. Make sure your mask is over your nose and mouth. Addressing the Court, you don't need to stand up. Just make sure you speak into the microphone and make sure that it's fairly close so the court reporter can hear you. OK.

All right. Let's see. Pulling up my agenda, let's look at the issues in ECF 47 and the letters that followed.

Defendant seeks to compel production of information regarding commercial performance of books published by plaintiff. I think I'm a little bit concerned that the request is so broad, and maybe you can show me otherwise or show me that you've tried to meet and confer and come up with something

that's more proportional. If anybody wants to talk about where the parties are at this point on this request.

MR. GRATZ: Absolutely, your Honor. So I agree with you that production of monthly commercial performance data for every book of the plaintiffs is not the ideal and most narrowly tailored way to go about doing this analysis. We would like, as the defendants, to do it a different way. We would like to do it by comparing to comparable books, and in meet-and-confer that's what we proposed, that we try and reach a set of mutually agreeable comparable books and have the data produced for those. Plaintiffs weren't interested in talking about reaching an agreeable set of comparable books, based on their view that no book is comparable with any other book for these purposes, notwithstanding the fact that, based on what we've learned in discovery, the plaintiffs compare books to the performance of other books that they consider comparable all the time for forecasting purposes.

So in order to be able to develop a view of what the comparable books may be, we think we may need all of it. Again, we don't think that's the most narrowly tailored way to do it, but it is --

THE COURT: I don't think that's proportional. So go on.

MR. GRATZ: We don't think that's the most narrowly tailored way to do it. We would like to do it in a narrower

way. We tried to meet and confer and find something narrower, and plaintiffs weren't interested in trying to find some middle ground on what books may be comparable.

THE COURT: Before I let Ms. McNamara speak, I'm admittedly not as well versed in the area of the law. I assume that you all are experts in this area of the law compared to me. But this does sound to me a lot like some of the disputes I hear in employment cases where one is trying to find what the comparators are, right, that -- what is similarly situated in terms of the employee -- plaintiff employee and other -- when a plaintiff is alleging, say, discrimination on the basis of sex or race or something else, what is the appropriate comparator, right? This seems like it's the same general concept, and we talk a lot about what's proportional.

So I guess now I want to ask Ms. McNamara to explain to me what your thoughts are on this dispute.

MS. McNAMARA: Thank you, your Honor.

Let me begin by underscoring I think you're correct in saying what's being demanded here is just extraordinary and clearly not proportional, but we need to take a step back, I think, initially, and focus on the fact that the rationale for this demand, however you establish the comparables, is that they want to establish evidence that they believe might support market harm or lack of market harm, and there's two means of market harm that are primarily at issue in this litigation.

One is that Internet Archive is not paying the customary license fee that actual libraries versus the Internet Archive paid to acquire and distribute e-books. And we've established, and there's no dispute, that there's a thriving market for licensing e-books to lending libraries. Internet Archive just refuses to pay that customary license fee. This demand has nothing to do with that form of economic harm. That is established, and I think the lens in looking on this and to further underscore the lack of proportionality is the fact that in the case they cited in support of this motion on *Davis v. The Gap*, there's a reference to *Ringgold v. BET*, which is a Second Circuit decision, where the court observed that the fact that an infringement had little likelihood of adversely affecting sales deserves little weight in the fair use analysis against a plaintiff alleging appropriation without payment of a customary license fee.

So that's the lens we need to look at this is that the lost sales, which is the only thing they're focusing on with this demand, in this context where we have a customary license fee established in the market is given little weight in the fair use analysis. So with that predicate, we can step back, and I think what's important here is that the plaintiffs have already produced a massive amount of financial data concerning the 127 representative works at issue in this litigation. That data includes more than 670,000 lines in an Excel spreadsheet.

That data allows them to provide the comparison that they're seeking. That data, for the most part, while it varies from book to book but goes back as far back as 1996 for some of the works, they can run a comparison of sales of those books prior to the time they appeared on the Internet archive and after the time they appeared on the Internet Archive and make whatever argument they want to make about lost sales or lack of lost sales. And that is a comparison that is an apples-to-apples comparison with the same books and the sales and pre- and post-times on Internet Archive.

What they are proposing here is finding other books that are supposedly comparable. They want, in effect, to make an apples-to-zebras comparison. Mr. Gratz is correct when he says that we have objected to identifying comparable books for purposes of this type of comparison, because we can't do that, nor do we really believe that you can, in fact, do that in any coherent, concrete way. These works, like, for example, the works-in-suit include *Song of Solomon* by Toni Morrison, Sylvia Plath, *The Bell Jar*, Salinger's *Catcher in the Rye*. What other novels, whether they had comparable sales or not, would be deemed actually comparable to those books?

And no one can dispute, nor do they, that book sales rise and fall based on myriad and random events. An author dies, often the sales go up. An author gets arrested, the sales go down. They appear on Oprah, they go up. The book is

made into a movie, there's myriad ways that sales are dictated, and so being able to pick out of this choice and finding actual coherent comparables makes no sense.

Now, the testimony that he's referencing from certain witnesses, they have described and made it crystal clear that it's an art, not a science. That when they sometimes are determining -- when they're trying to determine what advance to pay an author, they sometimes run comparisons to kind of just make a gist. But it's like a back-of-the-envelope issue that really doesn't have any kind of scientific proof, and what they're trying to establish is evidence to put into a court record showing comparable sales and an increase or loss.

Finally, I'll note, your Honor, on the burden here, the burden here is beyond extraordinary, as your Honor has already recognized. This literally would take months in order to gather. There are multiple databases that the data would have to be drawn from. There will be literally billions of lines of data concerning these works. We're talking about a half a million books published by these four publishers in the ten-year period. The burden would just be extraordinary. We don't know how to begin to find the comparables.

I end with underscoring, again, they have the data to run comparables. They have all this massive data, and we're at a loss on why that isn't sufficient.

THE COURT: All right. What I'm hearing, and for my

simple mind, is that you're also making a relevance argument and also that the data is not -- that the information that you would use to support sales calculations pre and post a book's appearance in the Internet Archive has already been provided?

MS. McNAMARA: Correct, your Honor.

MR. GRATZ: May I address those, your Honor?

THE COURT: Sure.

MR. GRATZ: As to relevance, there isn't any comparable license for what the Internet Archive is doing, that is, owning a physical copy of a book, digitizing it, and lending it to one patron per owned physical copy. So you can't just look to the license fee and say nothing else is important; that is, if what Internet Archive did didn't affect at all the revenues either from print sales, e-book sales, or library lending licenses, that that just doesn't matter to the question that the statute asks, the question being to what extent did this have an effect on the market for or value of the copyrighted work?

So as to relevance, we want to be able to say, and we think we will be able to say, they didn't lose a dime. They didn't lose one licensing opportunity. They made 100 percent of their value; that is, this wasn't replacing transactions that otherwise would have happened in a way that got them money. We think we can do that by showing that compared to things that this wasn't happening to that are comparable,

there's no difference.

As to --

THE COURT: Wait, wait. How does Internet Archive make its money?

MR. GRATZ: Internet Archive does not make its money. It's a 501(c)(3) public charity. Internet Archive operates on grants, including government grants that support what it does in the public interest.

THE COURT: All right. The other thing I'm hearing -- I know I cut you off on the relevance argument, but I'm also concerned about the burden. Even if you got this information, if you're talking about billions of lines, of 500,000 books, how would you even analyze that?

MR. GRATZ: Let me respond on burden and then answer -- let me actually start by answering your Honor's question and then move to burden.

We would analyze it using statistical software. What we want to do is look for books that had sales -- sales graphs or things about their sales that looked pretty similar, looking for, effectively, twins that are out there before they were included in controlled digital lending through Internet Archive and then see whether at the time that changed, the time they started getting offered for controlled digital lending through Internet Archive, whether one twin diverged from the other. And we can't do that analysis, we can't find those twins,

without the whole set of the population. We're happy to try doing it some other way, that is, taking a random sample, seeing whether -- trying to find the twins some other way, taking a random sample of 127 other books and see how they match up. That, I think, your Honor, is my answer on burden.

One answer on burden, your Honor, is their witnesses have testified that this data's just sitting in a database. They can just pull it up; that is, they can just put in a book and get the data. Doing that on a mass basis may be harder. We're happy to talk about ways to lessen that burden. But this is not data that is in some inaccessible form. Their witnesses have testified that even the folks we are talking to are able just, on their computers, to pull up this granularity of data.

MS. McNAMARA: May I address that, your Honor, because that isn't actually accurate?

THE COURT: Just a moment.

MS. McNAMARA: Sure.

THE COURT: I guess before you address that, Ms. McNamara, I have just another question for Mr. Gratz.

Why can't you compare each book and the data from before and after that book appeared in the Internet Archive? Why is that not sufficient, or why is that not an appropriate first step at least?

MR. GRATZ: So that is one thing that we can do, your Honor. Certainly, that's something I think we're going to try

and do. One thing, by the way, that is hampering that is that they didn't produce monthly data even for the books-in-suit, which is another thing we're asking for in this motion.

But here's why we think we need to do more than that if we can, and that is, there's lots of extrinsic factors that affect the sales of lots of books. One of those, a really big one, is one that's central to this case: the COVID-19 pandemic. That is, comparing sales of a book in the past to sales of a book in the present doesn't necessarily tell the whole story without knowing something about the book market as a whole, at least for comparable books. So the concern, your Honor, with comparing a book to itself is that there may be confounding factors over time.

THE COURT: With that big fancy statistical analysis, you can do that analysis. I agree it might be harder if you don't have monthly data, and maybe we can explore that, but why can't you control for that factor in the analysis of pre and post and then see what you see?

MR. GRATZ: Well, I have a number of responses to that your Honor. One of them is we want to have the most robust analysis that we can. It is not that we don't think that would be a potentially valid way to do it. We're just thinking about what their expert's going to say in that expert --

THE COURT: What I'm saying here is let's take a measured -- let's not wholesale throw money at the problem

perspective of -- I'm trying to come up with an analogy, and I'm coming up short at this hour.

You don't necessarily need gold-plated Tiffany garden shears, if they even exist, to prune a bush. What I'm concerned about is that let's see if you can get some information out of what's already been produced and do that comparison, because you know what? If it turns out that you can make arguments from the comparison that you've already done that warrant going further, then I think you're in a much better position. But right now it just sounds like an almost inconceivably huge burden to put plaintiffs to to produce the information, and then also an exceedingly large burden on your part to analyze that data when you actually have a smaller data set, and maybe that data set could be expanded, but a smaller data set that you can do some preliminary analyses on and figure out whether this is an appropriate way to go forward.

The reason why I'm suggesting that is because otherwise I would be making you brief with case law the point that Ms. McNamara raised about what is actually an appropriate measure of damages, and I'm not quite sure we're there yet. We may get there in a month anyway, but I'd like to see what you can get with what you already have or some expansion on what you already have.

So, Ms. McNamara, you mentioned 670,000 lines on an Excel spreadsheet that relates to 127 representative works.

What's the form of that data? Because Mr. Gratz just said that that doesn't contain monthly sales.

MS. McNAMARA: Actually, I have a flash drive that has the data if your Honor wants to have it.

THE COURT: I shake my head for a number of reasons. One of which is our IT department would --

MS. McNAMARA: Kill you.

THE COURT: -- kill me. They have to go through their whole process to make sure that I don't take down the entire federal judiciary by plugging something into a computer here.

MS. McNAMARA: Of course.

THE COURT: And we've had concerns.

But the other part of it is if you can explain to me as a layperson, because you would have to do this or you'd have to have your expert do this to a jury later --

MS. McNAMARA: Yes.

THE COURT: -- or is this a bench trial? But anyway, you have to have your expert explain what this data is anyway.

MS. McNAMARA: Yes.

THE COURT: I'd rather not get into this briefing about what the appropriate measure of damages is this early in the case. So go ahead.

MS. McNAMARA: Thank you, your Honor. I'm happy to try to do that, although I probably suffer from some of the same problems/burdens you have with explaining all this.

But at any rate, what we have produced for each of the 127 works-in-suit is data that shows all the comparable factors that they've identified and that they want, the various distribution channels, the prices, the revenue streams broken down between paper, hard cover, e-books, and the related data. In order to do that, they had to pull that data. That's what I wanted to correct Mr. Gratz seemingly suggesting that this is like a button you can push. This data had to be aggregated from multiple databases and pulled together and pulled out over the -- dealings going back 10, 15, 20 years for some of these works, and it was an enormous undertaking even for the 127 works.

Now, some of the publishers, namely, specifically Hatchette, routinely have this data in monthly indications, and we have produced for all the Hatchette books monthly data. We've similarly produced monthly data for many of the Harper Collins works. The two publishers, Penguin Random House and Wiley are the ones that there has not been a production of monthly data. We would entertain some limited supplement on the monthly data in order for them, the defendant, to make this comparison that we think is readily available from this data. But I would underscore that we don't need monthly data going back from the first time this book appeared, like when Salinger first published in the -- whenever that was, in the '70s or whatever.

THE COURT: OK.

MS. McNAMARA: Because, really, the monthly data is only important, arguably, in my mind, for two purposes: One is what Mr. Gratz already identified, which is during the pandemic, the world kind of shifted, as we all know, within a month or two, and so for that period of time in 2019, during the early stage of the pandemic, it seems to me that there's a rational reason to have monthly data so you can see what changed. It was also during that period of time that the Internet Archive for three months instituted what they called the "National Emergency Library," during which they threw all their supposed rules to the wind and just made these books available to anyone anywhere anyhow across the board. So for that period of time, I think there's a logic to monthly sales reports, and we can endeavor to provide some of that.

The other rational, I think, reason would be if there's a change. The Internet Archive knows -- and I'm not sure whether it's clear in the data they have provided to us -- but when each of these 127 works were originally published on the Internet Archive and made available to the world for free, and so I think that they indicate in their letter that if that happened in March of whatever year, there might be a reason to kind of do a comparison around months concerning when they were initially posted. If Internet Archive wants to provide us with that data or that information for the 127 works, I can speak to

my clients about whether there's some accommodation to be made on the monthly sales figures.

THE COURT: Mr. Gratz, anything to add?

MR. GRATZ: Yes, your Honor. That sounds like a fine first step; that is, what we need is monthly -- in order to do what your Honor is suggesting, which we are happy to do, that is, try to boil the pot before we see whether we need to collect the ocean, we would be happy to see whether monthly data about the works-in-suit for some reasonable period before they first appeared in Internet Archive's lending library, which data we have produced, but we're happy to point out or produce in a more readily usable form.

We also, your Honor, would need monthly data about the book market as a whole, or at least about these publishers' sales as a whole, in order to try and pull out macro trends that may have been sort of versus the micro trends versus the effect, which we don't think exists, of what Internet Archive was doing.

MS. McNAMARA: Your Honor, on that point, obviously, I don't object to the two other points regarding the time periods I suggested, but this macro -- there's been an ongoing back and forth with the defendant in this action about what data is provided from their data sets concerning the books and how they've made these books available for free and our data, and the parties have consistently tried to keep this to data points

surrounding the works-in-suit, and they have limited our ability to get a lot of information from them beyond the works-in-suit. So I really object to trying to backdoor all of a sudden getting data concerning entire book sales for these entire publishers for all the book sales that they have. I don't see the logic of that. I don't see why it's relevant and what the macro necessity is.

MR. GRATZ: Your Honor, as to that last point, this is data that they already make publicly available. They just didn't want to produce it and offered to sell it to us at its normal retail price when --

THE COURT: Wait, wait, wait.

MS. McNAMARA: What are you talking about?

MR. GRATZ: The Association of American Publishers, your Honor, which the same counsel already represents, we had asked Association of American Publishers for macro sales data, which it compiles.

THE COURT: Are we talking about the 127 works or the --

MR. GRATZ: We are not, your Honor. And this is --

THE COURT: Can we -- go on, but --

MR. GRATZ: I'll focus down, your Honor. I am not talking about per-book numbers. I am talking about single industry-wide numbers on a monthly basis. And the reason for that, your Honor, is if book sales as a whole doubled in a

particular month or halved in a particular month, the fact that the sales of one of the works-in-suit doubled in a particular month or halved in a particular month, one would want to take that into account in the analysis.

Does that answer your Honor's question?

THE COURT: Yes. But, I mean, these are -- this is information that you may be able to use without demanding that it be produced in a certain format.

MR. GRATZ: I agree, your Honor.

THE COURT: Yes. You can do -- I say a lot of "you can do" here without actually knowing.

Wouldn't that be a factor that you could account for in an analysis based on publicly available information?

MR. GRATZ: We think that we could try, although we would like, to the extent possible, to avoid a dispute about which publicly available information about overall monthly book sales is the reliable kind, which is why we wanted to get it from Association of American Publishers. And they said we won't give it to you in discovery. You should buy it from us.

THE COURT: How much does it cost?

MR. GRATZ: It was like $16,000.

THE COURT: There was a part of me that, when Ms. McNamara was talking about the burden of production of some of the data that you wanted, that I thought about mentioning that a lot of the time the burden is because it will be costly

and time-consuming, and sometimes the party that wants the data could pay for it or pay for the cost of -- or share the cost or pay for the cost of collecting, aggregating, and maintaining that data.

MR. GRATZ: But we have no objection, your Honor, to paying the actual cost of the Association of American Publishers giving us a copy of the reports they have already created, to the extent there is such a cost, but we don't think that's the number that they have suggested to us.

THE COURT: Go ahead.

MR. ZEBRAK: Thank you, your Honor. I appreciate your allowing me to chime in on this.

So if I'm understanding Mr. Gratz correctly, the data he's referring to with AAP, I think he's mixing apples and oranges here again, or apples and zebras, to use Ms. McNamara's reference. AAP collects data across the industry for kind of industry-wide sales figures and reporting, but I believe what they've been interested in is library sales information, which is not something that they track or report on. I believe this has already been explained to them.

And just on the whole notion of monthly data, the other thing that sort of sticks out to us is that if they understand that in year Y they scanned and started distributing copies of a work, I don't know why they couldn't use the sales figure on a yearly basis for what they have in the following

years and do comparisons that way rather than sort of turning the companies upside down for this data.

Thank you, your Honor.

THE COURT: Normally, I like parties to try to work something out. What I heard that was some kind of quasi agreement was whether defendants might work with some of the information that they have from the 670,000-line Excel spreadsheet. Particularly, once I learned that at least two of the publishers have already provided monthly -- have provided monthly information, why can't you do that in the first instance and then see if you can make a better claim for why you need more? If you can't agree on a resolution to this discovery dispute, I'm going to have to ask you -- I'm going to have to order you to brief it, which is going to be expensive for both of you.

One of my favorite rules to cite is Rule 37(a)(5). So think about the cost to brief it, think with the potential cost if you were to lose a motion like that and how badly you really need this information as opposed to making the sufficient defense, because I think whether or not you get to this whole commercial performance of plaintiffs' books and finding the comparators is a question that any -- that remains in this case is Internet Archive was not paying the customary license fee paid to this -- that normally is paid to distribute these particular e-books.

What does that mean as far as liability and damages on whether plaintiffs are entitled to the license fee at all, right? Even if you were to get this information that you seek, and I'm not suggesting that you would, but even if you did and you put plaintiffs to all that trouble, does that have any bearing on what I'm hearing as defendant's failed to pay the customary license fee and that, at a minimum, is a measure of damages?

So think about that. I'm not going to rule on it, but I'm going to really suggest that we focus on the case and making sure that the case moves forward in an expedient and cost-sensitive way.

MS. McNAMARA: Thank you, your Honor.

MR. GRATZ: Thank you, your Honor.

THE COURT: Next issue is the issue on -- let me see, on the privilege, withholding on privileged or arguably privileged communications. I thought it was an ECF 47. I guess I'm not sure I'm seeing it.

MR. GRATZ: It is. If I may, your Honor, it's ECF 54, and the response is ECF 56.

THE COURT: I missed ECF 54, but I have ECF 56.

ECF 54, I guess, was plaintiffs, right?

MR. GRATZ: ECF 54 was defendants, your Honor.

THE COURT: All right. Why don't you tell me what the dispute is.

MR. GRATZ: Certainly, your Honor. There are two deeply intertwined disputes here. They both relate to communications among members of the Association of American Publishers, the trade group that instigated this lawsuit about the Internet Archive and about library lending. Some of those communications are likely to be privileged, communications about selection of counsel, communications about legal strategy, communications among the plaintiffs, things like that. Many of those communications we think are not privileged, that is, communications about the business aspects of how to think about what the Internet Archive is doing, how to think about what libraries in general are doing and how it affects the plaintiffs' businesses.

Many of the withheld communications with respect to -- that have been logged by the plaintiffs include people who are neither plaintiffs nor the AAP, and the plaintiffs are claiming that the AAP, the trade association, is acting as their lawyer in this context. While it's conceivable that a trade association can act as a lawyer for members under some circumstances, the case law sets forth a variety of showings that would need to be made in order for there to be a finding that a trade association, who isn't a lawyer but might employ lawyers, is acting as counsel for a particular member in a particular communication.

THE COURT: You already have a privilege log?

MR. GRATZ: So with respect to the plaintiffs', your Honor, we do. With respect to the AAP, they've refused to produce a log.

MS. McNAMARA: Your Honor, do you want me to address this?

THE COURT: Yes, please.

MS. McNAMARA: Thank you very much, your Honor.

THE COURT: You can bring it right up to your mouth. That's better for all of us.

MS. McNAMARA: OK. Great.

In reviewing the exchange of letters in preparation for this hearing today, I was struck by the fact that there is no issue that the Internet Archive is really presenting to this Court. There's nothing concrete that they have identified where they have a concern about the production, or lack of production, looking at actual documents that have been withheld. There's no dispute amongst the parties as to the legal principles at issue here. Mr. Gratz just identified some of them. No one disputes that some communications can be privileged between the AAP and its members. There's no dispute that some communications concerning lobbying, and the like, with Congress can be protected by a First Amendment privilege, and there's no dispute that a privilege doesn't arise just out of mere membership in an organization.

But the plaintiffs have not withheld any documents

based on that contention. Internet Archive doesn't point to a single example, doesn't identify for this Court a single example, that reflects that the plaintiffs did withhold documents simply because they were communications with AAP. Instead, it's clear that the privilege log makes clear, and the plaintiffs know this, that we have produced a number and large number of documents, of communications between AAP and the plaintiffs or others, including others when they dealt with business issues. No one disputes that you have to analyze documents as to whether this is rendering legal advice or concerning business issues. The documents have been carefully reviewed to withhold only those documents that dealt with privilege issues, and that's reflected on the privilege log, and the letter to your Honor doesn't identify a single example where it appears to be citing to business issues. Instead, the log would say something like "legal advice re anticipation of a copyright infringement litigation against IA." That is clearly legal advice.

Now, when this issue first arose and they said, well, you need to establish that there was a attorney-client relationship between the parties, that testimony has now been rendered in this case, that has been asked in depositions. There was the deposition of Skip Dye, who's one of the key witnesses here, and he had indicated -- was asked squarely:

Does the AAP serve as its attorney outside of this

lawsuit?

Yes.

Does it provide legal advice?

Yes.

And the testimony of Alison Lazarus from Hatchette -- Skip Dye is from PRH -- was asked similarly: Does the AAP act as Hatchette's counsel, and do they give legal advice from AAP? And the answer is: Yes, I believe we do.

Now, these are not -- you have to understand who the AAP is and who its executive director is in order to appreciate that this is bona fide, compelling legal advice. The executive director of the AAP is a woman by the name of Maria Pallante. She was the former register of copyrights and the director of the U.S. Copyright Office. She is far more equipped to give advice on copyright issues than I could begin to be, and even though I practice in the area of copyright quite a lot, she is far more of an expert than I.

THE COURT: Let me just pause, because it's getting late.

MS. McNAMARA: Yes.

THE COURT: I feel like an air traffic controller or maybe, yes, the LaGuardia during a thunderstorm.

What are the documents that defendants seek to compel, who produced them, and is there a log?

MR. GRATZ: So, sorry, is that a question?

THE COURT: Yes.

MR. GRATZ: So the answer is with respect to the documents in the possession of the publishers, we have not at this point in the dispute process teed up in our letter which log lines we think should be produced. We have identified them instead by category, for example, communications between -- solely between nonparty third parties and communications between plaintiffs and third parties who are not the AAP, and we think they need to redact a number of the documents that are likely to be mixed.

THE COURT: OK.

MR. GRATZ: Then with respect to the AAP, who we have subpoenaed, your Honor, they haven't produced a log.

THE COURT: AAP isn't here.

MR. GRATZ: They are here, your Honor.

MS. McNAMARA: Yes, AAP, but, your Honor, this is not the proper jurisdiction for AAP. They were subpoenaed in Washington, D.C., and I think that if there's any motion to compel with regard to the AAP, it needs to be held in D.C., not before your Honor.

THE COURT: So the AAP is in D.C. They're a nonparty.

MR. GRATZ: May I be heard on that?

THE COURT: No, not yet.

AAP is in D.C. They're a nonparty. AAP and AAP's docs are in D.C., is that right?

MS. McNAMARA: Correct, your Honor.

THE COURT: Let's go back to the documents in the publishers' productions.

Do you want to brief it? I'm not sure that you're ready to brief it. I think you probably need to meet and confer and have another round because what I heard was we're complaining about particular categories, and we think that things something should be produced in redacted form. I could put you on a briefing schedule or I could put you on a briefing schedule after you have had another chance to meet and confer on the documents produced by plaintiffs for which you have a log.

MR. GRATZ: We think, your Honor, in light of today's discussion, that it might make sense to see whether there is at least some common ground we can reach with respect to the things to which we have a log, with respect to at least the production of redacted documents where there is mixed business and --

THE COURT: How about this? Let's set a date for a status letter for you to tell me where, if any, dispute remains on this issue. I don't want to make people work over the holidays if they weren't already planning to. I don't even want to say "if they weren't already planning to." I don't want to be the cause of an attorney, particularly an attorney not in this room, having to make changes to their holiday plans

or giving up time spent with their family to deal with this issue.

MS. McNAMARA: Thank you, your Honor.

THE COURT: I'd like you to spend some time to meet and confer, get back -- is a status letter in January appropriate?

MS. McNAMARA: That's fine.

MR. ZEBRAK: That's fine, your Honor.

THE COURT: Mr. Gratz wants to say something. Now that you're not sitting next to each other because of this strange COVID arrangement, you can't see when Mr. Gratz wants to say something. Yes, Mr. Gratz.

MR. GRATZ: I will wait a moment for Mr. Zebrak to finish whatever Mr. Zebrak was saying.

What I was going to say was --

THE COURT: No, but you're not waiting because you just said you would wait and then you started talking.

Go ahead.

MR. GRATZ: I'm sorry, your Honor.

A status letter in January sounds fine. We currently have close of discovery in two weeks, and I think we need to do something about that.

THE COURT: Yes, I will extend that date.

Mr. Zebrak, what were you going to say?

MR. ZEBRAK: Thank you, your Honor.

So there's actually a third discovery dispute that the plaintiffs brought here.

THE COURT: I know. I'm just trying to --

MR. ZEBRAK: Very quickly.

THE COURT: -- get through.

MR. ZEBRAK: I think we could do it very quickly, your Honor.

THE COURT: No, but I want to close out this issue unless you're -- I want to set a date for a joint status letter.

MR. ZEBRAK: Yes, your Honor.

THE COURT: We close out one issue, and then we move to the next one, right?

What's the date for your joint status letter on this? Is mid-January OK, or would you like the end of January?

MS. McNAMARA: Mid-January's fine, your Honor.

THE COURT: January 14 status letter on the privilege log issues.

Although I hate to burden a sister court, my question is why couldn't you or wouldn't you bring any subpoena issues with AAP's production in D.C.?

MR. GRATZ: I think the answer, your Honor, is we could, but I think it would be exceptionally inefficient to spin up something else there. At the time we served the subpoena, the place for production was in D.C. since we didn't

know that Ms. McNamara would be representing with -- her office in New York would be representing the AAP in this matter. We can just serve a subpoena with the place of performance to be New York and solve it that way. Of course, Rule 45(f) also allows us to file it there and then consent to bring it here.

THE COURT: OK. Yes.

MR. ZEBRAK: Your Honor, the place of compliance is where the nonparty's located, it's not where counsel chooses to bring the document.

THE COURT: Right. My question is if this is going to end up being a motion to compel AAP, and I'm not sure that it's ripe yet because it doesn't sound like there's been production or a log, but I'm just trying to get a little bit ahead of this problem. Can you guys please agree on a place to deal with this motion, if you can't actually address it without having to file a motion?

MR. ZEBRAK: Yes, your Honor, we will endeavor to resolve this with defendant's counsel. The issue here is that it's an incredibly broad subpoena that they've made no effort to narrow.

THE COURT: I don't want to hear your argument about the subpoena. I understand you have a dispute about the subpoena. I'm just trying to see whether the dispute has -- if you can't resolve the dispute, whether it has to be brought here or whether you can agree to either have it brought here.

I'm happy to hear it if I have jurisdiction. I'm happy to hear it if you agree, and I would actually prefer that you find a way to agree. Whether it's to agree to bring it here or whether you agree to litigate it in D.C., I don't really care. I mean, obviously, it would be a little easier if I didn't have to deal with it. At the same time, since I'm dealing with the main case, it does make some sense for me to deal with it. So please work that out.

I think that does mean that -- meet and confer status letter on the logs of the documents that have already been produced is January 14. OK. I think that now goes to Mr. Zebrak's issue. That's ECF 58 and 59?

MR. ZEBRAK: Yes, your Honor. Thank you.

Several issues have been resolved, so it would help me to address --

THE COURT: Why don't we start out with the issues that have been resolved.

MR. ZEBRAK: Thank you, your Honor.

I can begin by saying that the numbers that I believe have been resolved are Categories 3, 5, 6, and 7. For Category 3, the defendant produced documents after we filed our letter. For Category 6, the defendant produced documents after we filed our letter. For Category 7, we haven't yet reviewed the documents, but we understand that last night the defendant produced documents. And for Category 5, the defendant has

indicated that the documents are inaccessible to them without undue burden. And at this time, I think we'll leave it at that. The defendant -- I can get into more detail, but I think we can just focus on the three that still remain outstanding.

MR. GRATZ: With respect to Category 5, we may actually turn out to have some additional documents for you that I will give you an update on as soon as I have it, Mr. Zebrak.

THE COURT: OK.

MR. ZEBRAK: Thank you.

THE COURT: All right. Let's look at Category 1. What is Category 1?

MR. ZEBRAK: Sure. Category 1, just one quick backdrop. The defendant distributes e-books. The defendant does it two ways. Some are e-books in the public domain that the defendant distributes without any restrictions; someone you can just download and keep a permanent copy forever. The other way the defendant distributes e-books, typically, e-books that the defendant has scanned from physical form into a digital version, the defendant does it from what it calls its lending library. That's what this focuses on. It concerns defendants' rules, policies, procedures, specifications for how it's supposed to work, and the actual computer code that designates something into the lending library. This is very significant because they distribute millions of copyrighted books that

they've digitized from a physical form. So their rules for how something makes its way in the lending library and what's associated with that are quite relevant in this case. It's at the center in a lot of ways.

We've been following up on this for quite some time, and what the defendant has done is indicate that there's a single document from July of 2018 that they've produced that supposedly lays out essentially their intent for what gets designated into different collections, including the lending library.

Your Honor, that just doesn't quite make sense that there would be one document only from 2018, from three years ago, when this is so central to a process that's been the subject of so much controversy, that there's not documents concerning the implementation of these rules, updates to these rules. It's a single two-page document from 2018 they've identified to us as being the only thing that they have that they've produced. We would like them to be ordered to produce whatever else exists that they have not yet collected and produced.

THE COURT: Mr. Gratz.

MR. GRATZ: Two observations, your Honor: One, we are not making a relevance or a burden or an anything else argument. We have looked for documents. We have produced what we were able to find. We have identified by Bates number the

key document. And I think, most relevantly, plaintiffs are taking the deposition in about a week and a half of the person responsible for this, the 30(b)(6) designee on this topic. They can ask about this topic. They can ask about what documents are there, and they can satisfy themselves that they understand this document, they understand the procedures. And to the extent they can get testimony about some document that, despite our diligence, we haven't found, we'll give it to them.

THE COURT: When's the 30(b)(6) depo again?

MR. GRATZ: The 17th, which is currently the last day of fact discovery.

THE COURT: Right. Yes, I know. I will extend it.

All right. So on Category 1, take your deposition, see if you can resolve this one with a meet-and-confer after the deposition, and then you can put that in a status letter.

MR. ZEBRAK: We will, your Honor. Of course, the difficulty is that a witness is not going to be able from memory to cite different documents, and documents others --

THE COURT: But it's a 30(b)(6) deposition, right? So what he says is binding the corporation, right?

MR. ZEBRAK: Well, OK.

THE COURT: I have had cases where one would expect certain documents to exist, and it turns out after discovery, some discovery into processes, that such documents never existed in the first place or they were destroyed before the

litigation ever began.

MR. ZEBRAK: True. There is one thing now that I know exists which they haven't produced, which is the code that actually moves stuff into the lending library and indicates where it's pulled from and the records it's pulling from for all. So that's just one I know about right now that they can produce but haven't.

THE COURT: Mr. Gratz, what about that?

MR. GRATZ: I think we have produced that.

MR. ZEBRAK: We'll confer with them afterwards.

THE COURT: Yes.

MR. ZEBRAK: I don't believe that they have, but we will confer.

THE COURT: Category 2.

MR. ZEBRAK: Category 2, your Honor, so again one quick second of backdrop, otherwise this will also seem like it's written in a foreign language.

What the defendant does when it starts with a physical book is that it digitizes that book and then distributes copies of it from its website. Except during this period of what -- for three months during 2020, what the defendant says it does is that it limits one user at a time to that digitized copy of the book. However, what the defendant wants to do in order to scale how many copies it can deliver at a time to the world is that it reaches out to libraries and does what it calls an

"overlap analysis" with a library where the Internet Archive compares Internet Archive's data about the Internet Archive digital holdings of books to whatever data the library has about what the library thinks is in its physical holdings. The Internet Archive calls this an "overlap analysis." In this overlap analysis, the Internet Archive has what it calls "direct matches," and it also has what's called "similar matches."

THE COURT: OK.

MR. ZEBRAK: And after this overlap analysis occurs, libraries that participate in the analysis with the Internet Archive, afterward the Internet Archive will then just sometimes on the basis of this data comparison -- it doesn't just scan the library books. Essentially, Internet Archive treats all books as fungible and it then increases the concurrent lending count based on whatever match it thinks happened. And we've asked for the policies and procedures for how this works, not just stray emails here and there, and we said, including how you implement, whether you're going to use just the direct match or a similar match, again, you're using this as the basis to distribute millions of copyrighted books. You couldn't imagine that there aren't policies or procedures for this.

The defendant, despite our request, has not identified which documents they've produced that supposedly speak to this.

What they've said in their letter to your Honor is that they've produced documents relating to the overlap analysis. They haven't said they produced the documents that we've asked for, the policies the procedures, how do you implement this? We have none of that, and that's what we're asking for your Honor to compel them to produce.

(Continued next page)

THE COURT: I mean again, I can direct Mr. Gratz to produce things, but if they don't exist, I have a slight concern.

MR. GRATZ: This is the same witness who is --

THE COURT: Okay, but before we get to the witness, do documents exist?

MR. GRATZ: So documents do exist that describe this process, the documents Mr. Zebrak has been talking about.

Are there policies or sort of formal procedural documents or the things that Mr. Zebrak would sort of I think prefer exist that we're withholding? No, not that we found.

THE COURT: Okay.

MR. ZEBRAK: Your Honor, where I get hung up is when they say not that they were found, because, for instance the person that handles the designation "The Lending Library," that's Category 1, that's a non-custodian. They chose not to make that witness a custodian.

So we're deposing different witnesses, but this individual, Mr. Freeland, might have no idea about the documents that are in Category 1, and we're not going to learn of that unless he educates himself on it.

And it's the same thing here in Category 2. Mr. Freeland is not an engineer, he has an engineer that works for him and others that have done these overlap analyses, and Mr. Freeland, when he's deposed, may not know of documents.

And this is key.

THE COURT: Again, you are making an argument about the sufficiency of the 30(b)(6) witness's prep before the deposition has happened. Number one, I hope that your 30(b)(6) notice topics are broad enough to cover, and I hope that Mr. Gratz doesn't impose any bad faith objections on this so that you can understand and satisfy yourselves whether documents actually exist and also whether the preparation was sufficient.

MR. ZEBRAK: Thank you, your Honor.

THE COURT: And like I said, sometimes documents don't exist, and there is -- if you get to trial, they're not going to get to use that and they're not going to get to use a new document that they suddenly discovered. You all have laid the basis for excluding that document, so you're not without any recourse if it turns out that such documents that you would expect to exist don't exist or no longer exist.

MR. ZEBRAK: Thank you, your Honor. Fully understood.

THE COURT: Same thing, January 14, '21 status on whether this is remains to be an issue.

Category 4, which I understand is our last category. Then we push the discovery end date and the people who have been waiting patiently in my courtroom for hours today, I will then talk to them.

MR. GRATZ: I hope that I can shortcut Category 4,

your Honor --

THE COURT: Okay.

MR. GRATZ: -- by saying I think we have some additional data that I think will satisfy this. I'm happy to explain a little bit, but we think it is work we are -- we have additional data that I expect us to be producing, frankly today, that will provide at least sort of some additional information on Category 4, and the witness who is to testify about Category 4 and how those records are kept, his deposition, as of now, is tomorrow.

THE COURT: Okay. Mr. Zebrak, response?

MR. ZEBRAK: Your Honor, I'm not sure what to do with that exactly because we have been trying to obtain an answer on this for really months, and --

THE COURT: Funny how an in-person court conference often shakes things loose.

MR. ZEBRAK: First of all, I don't know what they intend to produce the night before his deposition, but we have not been getting our questions answered about an issue that is something that they believe is core to their case. The defendant wants to argue that there's no harm here because the books are only borrowed and used for a short period of time. And that's something they trumpet again and again and again. And we said if you want to make that claim, we have to see the background for it. You can't go to trial on statistical and

other claims without providing background.

And they referred us in two directions; both have been, thus far, dead ends. For the works in suit they produced essentially two different types of files with technical data, your Honor. We are referring to them in file format just for shorthand, but one is a CSV file. And the CSV file, there is data there on loans, but you can't use it to see when one loan begins and one ends because they do this thing with concurrent multiple loans, and it gets in the weeds quickly, but we're all in agreement between the plaintiffs and the defendants that you can't see when one loan begins and when one loan ends looking at that data.

THE COURT: But I thought you said earlier that Internet Archives supposedly only loans one copy, like one user at a time.

MR. ZEBRAK: That's one model, and sometimes they will do concurrent users where if they believe there's a library that also has a physical copy --

THE COURT: That's where you get to the overlap analysis.

MR. ZEBRAK: So without digitizing that book, they just said now it goes to two, three, four, then it's all thrown into these files in a suit and you can't see where one begins and one ends. That's the CSV files.

We then said OK, there's these JSON files that you

produced but there's a lot of essentially encrypted or scrambled or obfuscated data. We said: What does that mean? Can you produce it in a way that will allow us to see when loans begin and end? So we contest your narrative. And we have never gotten an answer to that question, including in response to their letter. So that's the technical data.

And then the other place they referred us, they referred us to two places, the technical data, and then they referred us to a bunch of handwritten analysis about how long books were supposedly used, I think it was on a given day, I don't recall the specifics right now, but we said so there's a lot of summary level data and analysis there by whoever wrote this by hand, give us the back-up data. You can't just give your summary without that backup.

And again, they didn't respond to it here, and our basic proposition is if you want to make claims about how long books are borrowed, you got to produce the data to support it. And that's essentially what we're looking to see compelled here, and that's Category 4.

THE COURT: Okay. Mr. Gratz, briefly, please.

MR. GRATZ: We have no objection to producing what we have that is reasonably accessible. As I said, there is some additional data on this that I think ought to be satisfactory -- although whether it will be, I don't know -- that has just shaken loose that will allow you to get some

breakdowns of buckets of how many loans were how long in certain period of time.

There is other more granular data that is truly stupendously difficult to get at, which is in the form of what we call web logs. It's just so voluminous that it's hard to get at. We're happy to talk about sampling or something to lessen that burden. That is a discussion that's ongoing that I hope, in light of what I hope we produce today, that it won't even be a discussion we need to have.

THE COURT: Okay. Again, let's put this in the January 14 status also, whether you resolved it or whether a dispute remains.

I wrote down summary analysis with handwriting, maybe that's an appropriate witness for deposition if there aren't documents.

Mr. Zebrak?

MR. ZEBRAK: It will be helpful to know what data he's producing tonight.

MS. McNAMARA: If this is for the witness tomorrow, we might need to have a supplemental deposition of that witness.

THE COURT: I expect that you would work together to make that happen.

MR. GRATZ: Or kick the deposition if we need to.

THE COURT: It's scheduled for tomorrow. They have been preparing. The deposition is going to go forward. I

don't think it's really appropriate to use last-minute produced documents to adjourn the deposition.

MR. GRATZ: We are happy to proceed either way.

THE COURT: Please do, because if it turns out that because of some of these document issues that there might need to be additional time allocated to a certain witness's deposition, again, you know how to write letters to the Court and ask for things, so we'll deal with that if and when it becomes ripe.

Let's set a fact discovery end date now of January 31. I don't expect -- I start out the day supremely optimistic, and by the time I get to the end of each day my optimism gets tarnished and wanes and sometimes falls into the gutter, but let's set a discovery end date of January 31, 2022, which perhaps, if you were able to resolve all of the disputes, and you were able to say so on January 14, that might be a realistic end date, but I won't hold you to it because I understand there are a lot of issues outstanding. But let's keep that as a placeholder so that come January 14 I will see what remains.

Let's make that a wholesale discovery status letter also, because then at that point if you do need more time, you will be able to tell me with some specificity what is still outstanding, what disputes remain, and sort of where along the ripeness scale they are. And probably what I will do is have

another in-person conference, but we'll schedule it after the January 14 letter. Okay?

MR. ZEBRAK: Thank you very much, your Honor.

THE COURT: Anything else that we need to do at this time?

MR. GRATZ: No, your Honor.

THE COURT: Have a happy and safe and healthy holiday. We are adjourned. Please order a copy of the transcript, split it 50/50.

(Adjourned)